Algorithm and code for analyzing hyperspectral images using the Hurst exponent

Nicolae APOSTOLESCU*[,1], Daniela BARAN[2]

*Corresponding author
*[,1]AEROSPACE Consulting
B-dul Iuliu Maniu 220, Bucharest 061126, Romania
apostolescu.nicolae@incas.ro, apostol@incas.ro
[2]INCAS – National Institute for Aerospace Research "Elie Carafoli"
B-dul Iuliu Maniu 220, Bucharest 061126, Romania
baran.daniela@incas.ro, dbaran@incas.ro

DOI: 10.13111/2066-8201.2014.6.3.1

Abstract: *The main goal of this paper is to present the implementation of an algorithm developed to calculate the Hurst exponent (H), applied here to characterize the pixel spectrum from hyperspectral image. Because a hyperspectral reflectance curve from each pixel may be regarded as a chaotic series that fact inspires us to treat a spectrum as a time series. Hyperspectral data are typical heteroscedastic variables, which makes it inappropriate to apply the normal/classic time series analysis, such as the autoregressive-integrated-moving-average model. Generally, the Hurst exponent is a measure used in nonlinear time series analysis to reveal local trend of series among adjacent successive terms. H can describe the local change in the ratio between the ranges of accumulated mean-removed values to the original standard deviation and thus represents the diversity of spectral values. Although H may be used to characterize regions of the image regarding persistence or antipersistence spectrum (highlighting noisy data), it does not directly address the separation between the classes of interest. The algorithm uses the rescaled range analysis method. This method introduces a measure of the variability of a time series using a ratio range/standard deviation (R/S). The algorithm was tested on hyperspectral data with spectra of various lengths and with persistence or antipersistence spectrum.*

Key Words: *hyperspectral data, spectral profiles, time series, Hurst exponent, persistent spectrum.*

1. INTRODUCTION

Hyperspectral images are important source of information which is used in many environmental assessments and monitoring of agriculture, meteorology, mineralogy etc. These data obviously provide much more detailed information about the scene than a normal color camera, which only acquires three different spectral channels corresponding to the visual primary colors red, green and blue.

Hyperspectral data sets are generally composed of about 100 to 200 spectral bands of relatively narrow bandwidths (5-10 nm), whereas, multispectral data sets are usually composed of about 5 to 10 bands of relatively large bandwidths (70-400 nm). Hence, hyperspectral imaging leads to a vastly improved ability to classify the objects in the scene based on their spectral properties.

Due to the rich information content in hyperspectral images, they are uniquely well suited for automated image processing, whether it is for online industrial monitoring or for remote sensing. Efficient exploitation of hyperspectral images is of great importance in remote sensing.

Hyperspectral images contain abundant spatial, spectral, and radiometric information of earth surfaces, which makes earth observation and information acquisition much more effective and efficient for material applications.

Images acquired from hyperspectral sensors contain many more bands and potentially more information than multispectral images, but hyperspectral images also tend to contain more noise, especially when acquired using small aircraft.

Hyperspectral data are spectrally overestimated and the useful signals usually occupy lower dimensional subspace which needs to be inferred. Therefore it is necessary to explore the dimensionality reduction methods which can effectively reduce noise in data sets with minimum loss of information. The signal information is usually concentrated in lower dimensional subspaces.

A hyperspectral pixel is generally a mixture of different materials present in the pixel data with various abundance fractions.

These materials absorb or reflect within each spectral band. As a consequence, spectral band characterization becomes crucial in hyperspectral analysis.

Since each pixel is composed of hundreds of spectral bands, the spectral information provided by pixel is generally very valuable in material detection, discrimination, and identification.

The paper is structured as follows: Section 2 contains an overview of the Hurst method, in Section 3 the Hurst algorithm is detalied while Section 4 gives the results on a real hyperspectral image.

2. HURST METHOD

Time and space series analysis methods have become widespread and valuable tools in studying hyperspectral data. A hyperspectral reflectance curve from each pixel is regarded as a chaotic series. This inspires us to treat a spectrum as a time series.

These data are typical heteroscedastic variables, which makes it inappropriate to apply the normal/classic time series analysis, such as the autoregressive-integrated-moving-average model. The Hurst exponent (*H*) is a measure used in nonlinear time series analysis. It is a classical value to detect long memory in time series.

The Hurst exponent's namesake, Harold Edwin Hurst (1880-1978), was a British hydrologist who researched reservoir capacity along the Nile River. He introduces as measure of the variability of a time series the statistical rescaled range.

H can describe the local change in the ratio between the ranges of accumulated mean-removed values to the original standard deviation and thus represent the diversity of spectral values. *H is* not limited to defined applications, and does not require training data, but it does not directly address the separation between the classes of interest into hyperspectral data.

The presence of large inhomogeneities in the series, such as large-magnitude abrupt changes in series variable (jumps or spikes), may lead to spurious results in detecting fractal scaling and calculating Hurst exponents [12].

The Hurst method provides an assessment of variability values from hyperspectral curve. Estimating the Hurst exponent for a data set provides a measure of whether the data is a pure white noise random process or has underlying trends. It is also used in characterizing stochastic processes. Using the Hurst exponent we can classify time series into types and gain some insight into their dynamics.

Practically, the Hurst exponent is a measure of autocorrelation (persistence and long memory).

- A value of $H \in (0,\ 0.5)$ indicates a time series with negative autocorrelation (e.g. a decrease between values will probably be followed by an increase). The series is antipersistent.
- A value of $H \in (0.5,\ 1)$ indicates a time series with positive autocorrelation (e.g. an increase between values will probably be followed by another increase). The series is persistent.
- A value of $H = 0.5$ indicates a “true random walk”, where it is equally likely that a decrease or an increase will follow from any particular value (e.g. the time series has no memory of previous values). Series of this kind are hard to predict because there is no correlation between the observations and a future observation; being higher or lower than the current observation are equally likely.

Also, the Hurst exponent is referred to as the “index of dependence” or “index of long-range dependence”.

The Hurst exponent is not so much calculated as it is estimated. A variety of techniques exist for estimating the Hurst exponent (H) and the process detailed here is both simple and highly data intensive.

To estimate the Hurst exponent one must regress the rescaled range on the time span of observations.

To do this, a time series of full length is divided into a number of shorter time series and the rescaled range is calculated for each of the smaller time series. A minimum length of eight is usually chosen for the length of the smallest time series.

Below we describe the algorithm from “Using Hurst and Lyapunov Exponent for Hyperspectral Image Feature Extraction” – Jihao Yin, Member IEEE, Chao Gao, and Xiuping Jia, Senior Member IEEE).

This algorithm uses the rescaled range analysis method. This method computes a ratio R/S (range/ standard deviation).

Rescaled range analysis reveals whether or not a timeseries exhibits persistence or anti-persistence bias.

R/S provides a simple tool for analyzing the time series in form of a plot. The Hurst exponent H, which ranges between 0 and 1, can be derived as the slope in the plot, in which ln(R/S) is plotted against ln(t), where t is the time step.

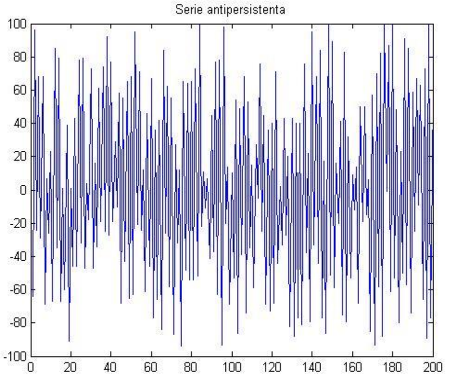

Fig. 1. An antipersistent series with 200 terms. H=0.3881

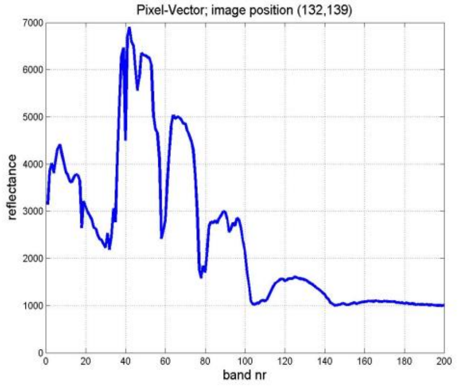

Fig. 2. A persistent series with 200 terms. H=0.9343

3. HURST ALGORITHM

A hyperspectral image can be illustrated as an image cube with the two dimensions of the cube face representing the spatial information and the third dimension representing the spectral information.

The spectral information available in a hyperspectral image (cube) is organized in a tridimensional matrix (each plane is an image corresponding to one wavelength band) named HS3D.

From the point of view of subsequent calculations we abandon the idea that the three-dimensional matrix represents the image; it is considered now that there is just a three-dimensional matrix that will be reorganized in a bidimensional matrix, HS2D, by the rule: each column of the bidimensional matrix contains the entire image from one spectral band.

Thus, the number of columns of the bidimensional matrix will be equal to the spectral bands number, and the number of the rows will be equal to the image dimension in one band.

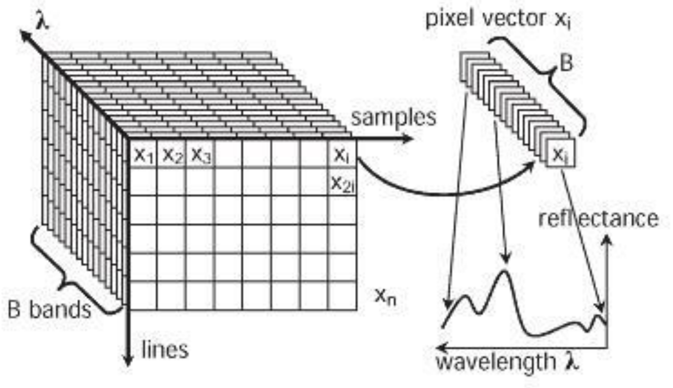

Fig 3. A hyperspectral image (from Yuliya Tarabalka, Jón Atli Benediktsson, Jocelyn Chanussot, IEEE, and James C. Tilton; Multiple Spectral–Spatial Classification Approach for Hyperspectral Data)

The vector $X_{ij} = (x_{ij1}, x_{ij2}, \cdots, x_{ijp}), \quad i = 1, m; \; j = 1, n$ is named, in this context, "spectral vector" of (i, j) pixel or "pixel vector". Pixel vectors or spectra are defined as the vectors formed of pixel intensities from the same location, across the bands. Each pixel vector is treated as a time series $x(x_1, x_2, \cdots, x_p)$.

Steps for estimating the Hurst exponent after breaking the time series into subseries:

For each subseries of observations, Hurst algorithm computes:

- the mean of the time series,
- a mean centered series by subtracting the mean from the series,
- the cumulative deviation of the series from the mean by summing up the mean centered values,
- the Range (R), which is the difference between the maximum value of the cumulative deviation and the minimum value of the cumulative deviation,
- the standard deviation (S) of the mean centered values, and
- the rescaled range by dividing the Range by the standard deviation. A high standard deviation indicates that the data values are spread out over a large range of values.

1. Divide the given series $\{x_k \mid k = 1,2,...,K\}$ into *D* subseries, where *K* is the total number of bands. The *d*th subseries is denoted as

$$X_d = \{x_k \mid k = (d-1) * n + 1, ..., d * n\}, d = 1,2,...,D.$$

The length of each subseries n is set between $n \in [\ln K, \sqrt{K}]$. Note the subseries elements with

$$g_{d,i} = x_{(d-1)*n+i}, \quad d = 1,2,\ldots,D;\ i = 1,2,\ldots,n.$$

2. For each subseries, we compute the mean m_d and the standard deviation S_d

$$m_d = 1/n * \sum_{i=1}^{n} g_{d,i},$$

$$S_d = \sqrt{1/n * \sum_{i=1}^{n} (g_{d,i} - m_d)^2}.$$

3. Normalize each subseries by subtracting its sample mean:

$$g_{d,i} = g_{d,i} - m_d,\ i = 1,2,\ldots,n. \text{ (n subserie' s length).}$$

4. Find the mean-removed cumulative series $\{c_{d,i}\}$ and its range Rd for each subseries $\{c_{d,i}\}$

$$c_{d,i} = \sum_{j=1}^{i} g_{d,j},\ i = 1,2,\ldots,n.$$

$$R_d = \max(c_{d,i}) - \min(c_{d,i}),\ i = 1,2,\ldots,n.$$

5. Calculate the mean value $(R/S)_n$ for all subseries with length n

$$(R/S)_n = \frac{1}{D}\sum_{d=1}^{D} R_d / S_d.$$

6. Repeat the aforementioned steps for all possible n.
7. The H is estimated by the following empirical equation.

$$(R/S)_n = c * n^H$$

By logarithm of both terms of the above equation, the H value can be obtained considering the expresion

$$\ln(R/S)_n = \ln c + H * \ln n$$

as a linear regression in which c and H can be estimated (using Matlab's polyfit).

4. EXPERIMENTS

Based on this algorithm we have built a matlab function to determine the Hurst exponent for each hyperspectral curve from a scene gathered by AVIRIS senzor over Indian Pines site in North-western Indiana. This hyperspectral image consists of 145X145 pixels in 200 spectral bands in the wavelength range 0.4-2.5 μm, with approximately 10 nm spectral resolution and 30 m spatial resolution. A hyperspectral image can be illustrated as an image cube with the two dimensions of the face of the cube represents the spatial information and the third dimensional representing the spectral information. The spectral information available in a hyperspectral image (cube) is organized in a tridimensional matrix (each plane is an image corresponding to one wavelength band) HS3D. From the point of view of subsequent calculations we abandon the idea that the three-dimensional matrix represents the image; it is considered now that there is just a three-dimensional matrix that will be reorganized in a bidimensional matrix, HS2D, by the rule: first column of bidimensional matrix includes all the columns of HS3D, placed successively, where, p=1; second column will contain all columns of the hs3d matrix, placed successively, where p=2, and so on. Each row can then be written as vector and regarded as a chaotic series. If the hyperspectral image consists of hXw pixels in p spectral bands, the 3D hyperspectral image will be HS3D(h,w,p) and the bidimensional HS2D(h*w,p).

The pixel $x(i, j, k)$ from (i, j, k) position in HS3D, in HS2D will be in $(k-1)hw + h(j-1) + i$ position, where $i = 1,2,\cdots,h;$ $j = 1,2,\cdots,w;$ $k = 1,2,\cdots,p.$

Using statistical language, we consider that the vectors pixels of a hyperspectral image are observations and their components are variable associated features.

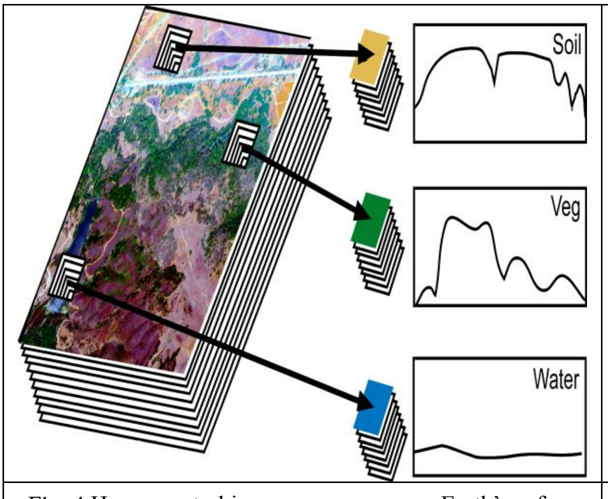

Fig. 4 Hyperspectral images on a scene on Earth' surface

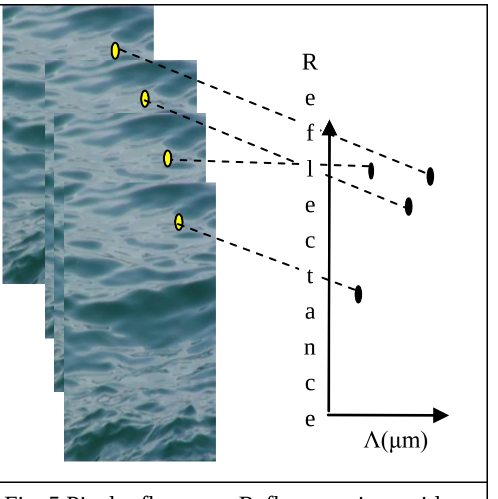

Fig. 5 Pixel reflectance. Reflectance is a unitless quantity that ranges in value from 0 to 1.0

These figures are from "Hyper&Multispectral _Imaging.pdf", Dr. Richard Gomez, Applied Technology Institute (ATI).

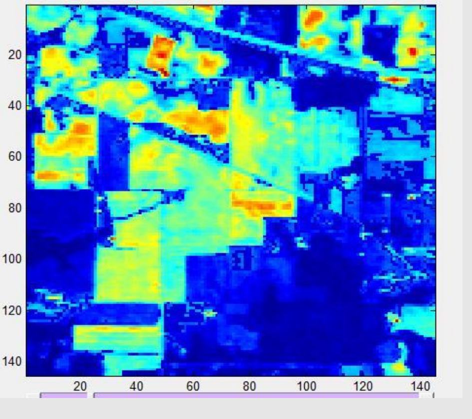

Fig. 6 Airborne Visible IR Imaging Spectrometer image of Purdue Indiana Indian Pine test site

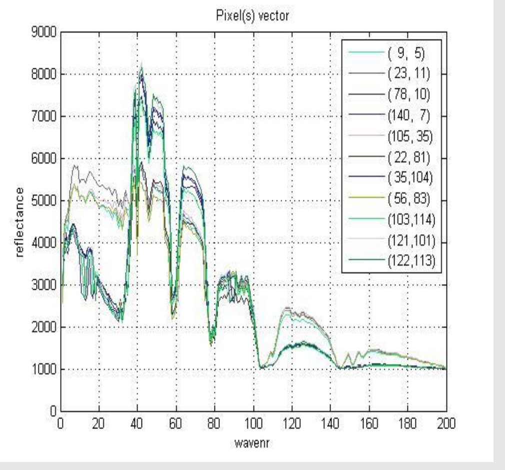

Fig 7. Pixel's vector (are highlighted their position in the original image)

H is the slope of regression line

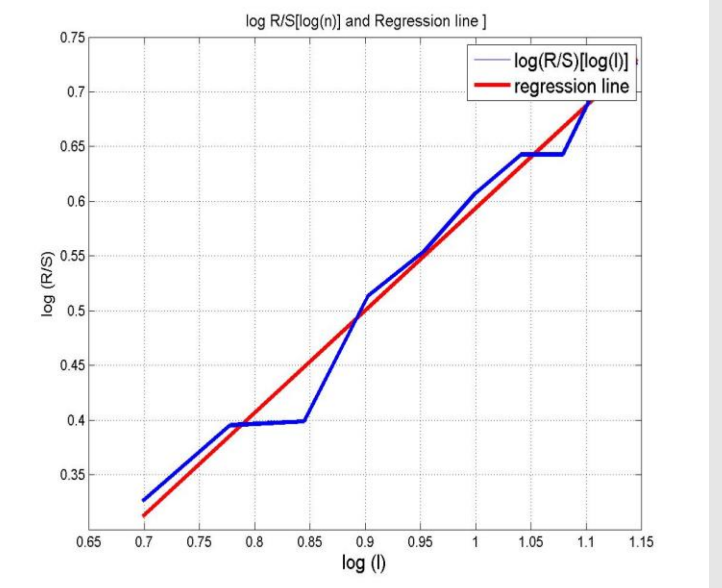

Fig. 8 R/S and regression line (H= slope of regression line)

0.9506	0.9222	0.9085	0.9348	0.9379	0.9455	0.9491	0.9314
0.9243	0.9175	0.9255	0.9523	0.9421	0.9536	0.9640	0.9599
0.9327	0.9406	0.9167	0.9373	0.9656	0.9278	0.9450	0.9356
0.9433	0.9125	0.9144	0.9220	0.9709	0.9513	0.9183	0.9501
0.9296	0.9171	0.9352	0.9319	0.9495	0.9609	0.9456	0.9594
0.9367	0.9425	0.9415	0.9248	0.9328	0.9713	0.9270	0.9304
0.9460	0.9593	0.9384	0.9415	0.9575	0.9562	0.9366	0.9753
0.9480	0.9548	0.9471	0.9297	0.9671	0.9415	0.9360	0.9423

Fig. 9 The H values for a small region of Indian Pine hyperspectral image (8x8 pixels - upper left corner). H>0.5 each pixel vector represent a persistent series

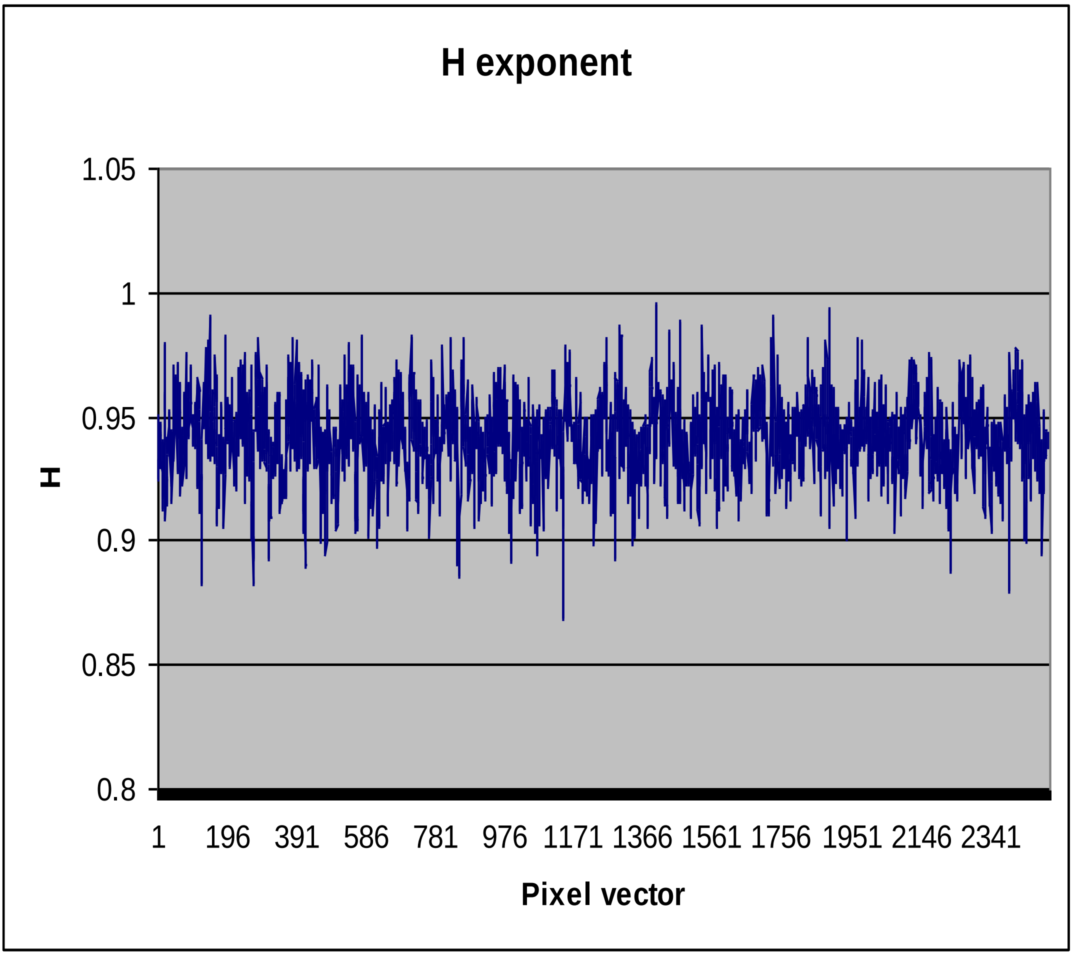

Fig. 10 The H exponent for a small region (50x50 pixels) of Indian Pine hyperspectral image. Because H>0.5 each pixel vector represents a persistent series

5. CONCLUSION

The Hurst exponent is a useful statistical method for inferring the properties of a time series without making assumptions about stationary. It is most useful when used in conjunction with other techniques, and has been applied in a wide range of industries. The Hurst exponent is frequently calculated for experimentally obtained data sets to characterize noise data series or in characterizing stochastic process.

REFERENCES

[1] J. Yin, C. Gao and J. Sun (School of Astronautics, Beihang University, Beijing 100191, China), *Verification of statistical properties for hyperspectral images: heteroscedasticity and non-stationarity*, The Third International Conference on Digital Information Processing and Communications (ICDIPC2013) - United Arab Emirates, ISBN: 978-0-9853483-3-5 ©2013 SDIWC, pp. 191-195.

[2] R. Racine, Estimating the Hurst Exponent, MOSAIC Group, Prof. Ivo F. Sbalzarini, ETH Zurich, Bachelor Thesis, April 14, 2011.

[3] J. Yin, C. Gao and X. Jia, *Using Hurst and Lyapunov Exponent For Hyperspectral Image Feature Extraction*; IEEE Geoscience and remote sensing letters, vol. **9**, no. 4, pp.705 – 709, July 2012.

[4] B. B. Mandelbrot, Multifractal Power Law Distributions: Negative and Critical Dimensions and Other ‘‘Anomalies’’, Explained by a Simple Example, *Journal of Statistical Physics*, Vol. **110**, Nos. 3–6, pp. 739 – 774, March 2003.

[5] B. Qian and R. Khaled, *Hurst Exponent And Financial Market Predictability*, IASTED conference on “Financial Engineering and Applications”(FEA 2004), pp. 203-209.

[6] J. A. O. Matos, S. M. A. Gama, H. J. Ruskin, A. Al Sharkasi, M. Crane, Time and scale Hurst exponent analysis for financial markets, *Physica A: Statistical Mechanics and its Applications*, Volume **387**, Issue 15, pp: 3910-3915, 15 June 2008.

[7] K. Cawse, A. Robin, M. Sears, *The effect of noise whitening on methods for determining the intrinsic dimension of a hyperspectral image*, DOI: 10.1109/WHISPERS.2011.6080974, in Workshop in Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011.

[8] Y. Tarabalka, J. A. Benediktsson, J. Chanussot and J. C. Tilton, *Multiple Spectral–Spatial Classification Approach for Hyperspectral Data*, IEEE Transactions On Geoscience And Remote Sensing, Vol. **48**, No. 11, pp. 4122 - 4132, NOVEMBER 2010

[9] T. Gneiting, M. Schlather, Stochastic models which separate fractal dimension and Hurst effect, *SIAM Review* **46** (2004), 269-282.

[10] K. E. Bassler, G. H. Gunaratne & J. L. McCauley, Markov Processes, Hurst Exponents, and Nonlinear Diffusion Equations, *Physica A: Statistical Mechanics and its Applications*, DOI: 10.1016/j.physa.2006.01.081, 2006.

[11] P. M. Robinson (editor), *Time Series with Long Memory*, Oxford University Press, 2002.

[12] S. Katsev, I. L’Heureux, Are Hurst exponents estimated from short or iregular time series meaningful?, *Computers & Geosciences* **29** (2003) 1085–1089.